Data Analysis Intern Task : User Behavior and Transaction Analysis Report (using SQL)

Name – Priya Chakradhari

Date – 04/11/2024

CONTENT

1. Introduction
2. Dataset Description
3. Objectives
4. Methodology
5. Analysis and Insights
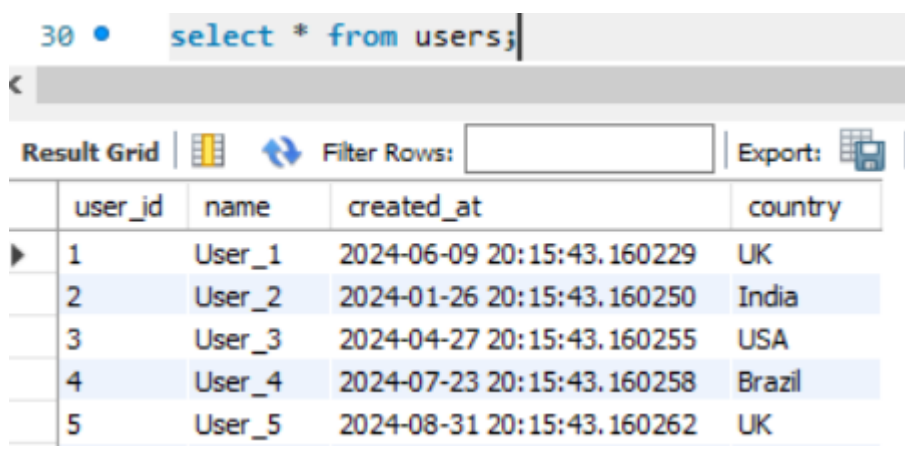6. Conclusion
7. Feature Scope

INTRODUCTION

This report presents an in-depth analysis of user engagement and transaction behaviors based on data from a platform's user, transaction, and session records. By exploring key metrics such as new user transaction rates, daily active users, average deposit amounts by country, and top spending users, this study aims to uncover patterns and insights that can inform strategic decisions. Key focuses of the analysis include understanding user engagement trends, financial behaviors by country, session duration by device type, and transactional frequencies. These insights provide a foundation for enhancing user experience, optimizing platform offerings, and driving targeted engagement initiatives.

DATASET DESCRIPTION

The dataset consists of three primary tables: users, transactions, and sessions. Each table provides essential data points that collectively offer a comprehensive view of user behavior and financial transactions.

Users: Contains information on each user's account, including a unique user ID, name, account creation date, and country.

- user_id (integer): Unique identifier for each user.
- name (text): Name of the user.
- created_at (timestamp): Date and time the user account was created.
- country (text): Country of the user's location.

Transactions: Contains information on individual transactions, capturing the transaction ID, user ID, amount, date, and type (deposit or withdrawal).

- transaction_id (integer): Unique identifier for each transaction.
- user_id (integer): Foreign key linking to the users table.
- amount (decimal): Amount of the transaction.
- transaction_date (timestamp): Date and time the transaction occurred.
- transaction_type (text): Type of transaction, either "deposit" or "withdrawal."

```
30 ●      select * from transactions;
```

transaction_id	user_id	amount	transaction_date	transaction_type
1	22	494.29	2024-10-31 20:15:43.162363	withdrawal
2	29	954.48	2024-10-17 20:15:43.162372	deposit
3	12	147.25	2024-10-20 20:15:43.162376	deposit
4	6	761.14	2024-10-27 20:15:43.162379	deposit
5	33	700.36	2024-10-09 20:15:43.162382	deposit

Sessions: Contains data on user sessions, including the session ID, user ID, start and end times, and device type (mobile or desktop).

- session_id (integer): Unique identifier for each session.
- user_id (integer): Foreign key linking to the users table.
- session_start (timestamp): Start time of the session.
- session_end (timestamp): End time of the session.
- device_type (text): Device used during the session, either "mobile" or "desktop."

```
30 ●      select * from sessions;
```

session_id	user_id	session_start	session_end	device_type
1	50	2024-10-20 18:15:43.163690	2024-10-20 19:22:43.163690	mobile
2	9	2024-10-27 08:15:43.163700	2024-10-27 08:29:43.163700	mobile
3	25	2024-10-09 23:15:43.163705	2024-10-10 01:06:43.163705	desktop
4	1	2024-10-30 22:15:43.163710	2024-10-31 00:12:43.163710	desktop
5	31	2024-10-07 21:15:43.163715	2024-10-07 21:39:43.163715	mobile

OBJECTIVES

The primary objectives of this analysis are to:

1. **Evaluate New User Engagement**: Determine the rate at which new users engage in financial transactions within their first 7 days on the platform, providing insights into initial user interest and onboarding effectiveness.
2. **Analyze Deposit Behavior by Country**: Calculate average deposit amounts by country and identify countries with substantial deposit activities. This helps in understanding regional spending patterns and tailoring strategies for different markets.
3. **Track Daily User Activity**: Measure daily active users (DAU) over the past 30 days to gain insights into user engagement patterns, identify peak usage periods, and assess overall platform activity levels.
4. **Identify High-Engagement Users by Device**: Find the top users with the longest session durations by device type, allowing for the analysis of user behavior on mobile vs. desktop and supporting device-specific optimizations.
5. **Determine Transaction Frequency**: Calculate the average number of transactions per user over the last 6 months, providing an understanding of transactional engagement levels across the user base.
6. **Highlight Top Spending Users by Country**: Identify the highest-spending users by country based on total deposit amounts, which can guide targeted marketing efforts and personalized engagement for high-value users.

METHODOLOGY

This analysis employs SQL queries to extract and process data for each objective from the user, transaction, and session tables. Each step focuses on a specific metric or pattern related to user engagement, transactions, or device usage:

1. **New User Transaction Rate:** Filtered users who registered in the past month and checked if they made any transaction within their first 7 days. Calculated the percentage of these users to gauge early engagement rates.

2. **Average Deposit Amount by Country:** Filtered transactions for deposits made in the last 3 months. Calculated the average deposit amount per country for countries with at least 5 depositing users, helping to pinpoint countries with significant transaction volumes.

3. **Daily Active Users (DAU):** Queried session records over the last 30 days to count the number of unique users with at least one session per day, producing a daily count of active users and showing trends in platform engagement.

4. **Top Users by Session Duration:** Aggregated session times for each user and device type, focusing on the last month. Ranked users by their total session duration to find the top 5 users for both mobile and desktop devices, indicating highly engaged users per platform.

5. **Transaction Frequency by User:** Filtered transactions within the last 6 months and calculated the average transaction count per user. This analysis reflects how often users make transactions, highlighting overall user interaction frequency.

6. **Top Spending Users by Country:** Summed total deposits for each user within the past 6 months and identified the top 3 users per country. This analysis focuses on high-value users, allowing the company to recognize top spenders across regions.

ANALYSIS and INSIGHTS

1. New User Transaction Rate

- Calculate the percentage of new users (users who registered within the last month) who made at least one transaction (deposit or withdrawal) within their first 7 days of account creation.
- Output the percentage as new_user_transaction_rate rounded to two decimal places.

```sql
select
    round(
        (select count(distinct t.user_id)
        from transactions t
        join users u
        on t.user_id=u.user_id
        where t.transaction_date <= u.created_at + interval 7 day and
        u.created_at >= now() - interval 1 month)*100.0
        /
        (select count(*) from users where created_at >= now() - interval 1 month),2) as new_user_transaction_rate;
```

Output

new_user_transaction_rate
66.67

2. Average Deposit Amount by Country

- Calculate the average deposit amount for each country over the last 3 months.
- Include country and average_deposit_amount rounded to two decimal places.
- Only include countries with at least 3 users who made deposits in this period.

```sql
select u.country,round(avg(t.amount),2) as avg_amount_deposit,
count(distinct u.user_id) as count_of_user_id
from transactions t join users u on t.user_id = u.user_id
where transaction_type = "deposit" and transaction_date >= now() - interval 3 month
group by u.country
having count(distinct u.user_id)>=3;
```

Output

country	avg_amount_deposit	count_of_user_id
Australia	582.73	3
France	589.19	3
India	657.65	3
USA	624.43	3

3. Daily Active Users

- Find the daily active users (DAU) for the last 30 days. Define DAU as users who had at least one session on a given day.
- Output date and daily_active_users (number of distinct users active on that date).

```sql
select date(session_start) as date ,
count(distinct user_id) as daily_active_users from sessions
where session_start>= now() - interval 30 day
group by date(session_start)
order by daily_active_users desc;
```

Output

date	daily_active_users
2024-10-22	5
2024-10-30	5
2024-10-07	4
2024-10-09	3

4. Top Users by Session Duration

- For each device type (mobile and desktop), identify the top 5 users with the highest total session duration (in minutes) over the last month.Include user_id, name, device_type, and total_session_duration

```sql
select s.user_id,u.name,s.device_type,
round(sum(timestampdiff(minute,s.session_start, s.session_end)),2)
as total_duration from sessions s
join users u on s.user_id = u.user_id
where s.session_start >= now() - interval 1 month and s.device_type = "Desktop"
group by s.user_id, u.name, s.device_type
order by s.device_type,total_duration desc limit 5 ;
```

Output

user_id	name	device_type	total_duration
48	User_48	desktop	299
32	User_32	desktop	173
10	User_10	desktop	125
1	User_1	desktop	117
24	User_24	desktop	111

For device type mobile

```sql
select s.user_id,u.name,s.device_type,
round(sum(timestampdiff(minute,s.session_start, s.session_end)),2)
as total_duration from sessions s
join users u on s.user_id = u.user_id
where s.session_start >= now() - interval 1 month and s.device_type = "Mobile"
group by s.user_id, u.name, s.device_type
order by s.device_type,total_duration desc limit 5 ;
```

Output

user_id	name	device_type	total_duration
45	User_45	mobile	204
13	User_13	mobile	199
3	User_3	mobile	190
16	User_16	mobile	162
28	User_28	mobile	154

5. Transaction Frequency by User

Calculate the average number of transactions per user within the last 6 months. Define this as the average count of transactions (deposit or withdrawal) for users who were active in this period.

Output a single value as average_transactions_per_user, rounded to two decimal places.

```sql
select * from transactions;
select round(avg(transaction_count),2) as avg_transaction_per_user
from (select user_id, count(*) as transaction_count from transactions
where transaction_date >= now() - interval 6 month
group by user_id) as user_transaction;
```

Output

avg_transaction_per_user
1.52

6.Top Spending Users by Country

For each country, find the top 3 users with the highest total transaction amount (only deposits) within the past 6 months. Include user_id, name, total_amount, and country in the result.

If there are fewer than 3 users for a country, include only the available users.

```sql
select * from transactions;
select t.user_id, u.name, u.country,sum(t.amount) as total_amount from transactions t
join users u on t.user_id = u.user_id
where t.transaction_type ="deposit" and transaction_date >=now() - interval 6 month
group by t.user_id, u.name, u.country
order by u.country, total_amount desc;
```

Output

user_id	name	country	total_amount
20	User_20	Australia	940.25
6	User_6	Australia	761.14
23	User_23	Australia	46.79
15	User_15	Brazil	638.23
14	User_14	Canada	1410.5
47	User_47	Canada	55.83
29	User_29	France	954.48
35	User_35	France	855.48
12	User_12	France	546.8
33	User_33	India	1219.81
43	User_43	India	714.22
30	User_30	India	696.58
48	User_48	Japan	1664.87
11	User_11	Japan	588.05
26	User_26	Mexico	268.83
1	User_1	UK	1502.76
45	User_45	UK	516.85
34	User_34	USA	888.79

Conclusion

This analysis provides valuable insights into user behavior, financial activity, and engagement patterns on the platform. Key findings include a snapshot of early user engagement rates, spending behaviors across different countries, daily user activity trends, device-based session preferences, transaction frequencies, and top spenders by country. These results highlight areas where user onboarding, platform engagement, and targeted marketing can be further optimized to drive retention and revenue growth.

Future Scope

1. **Enhanced Segmentation**: Further breakdown of user segments by demographics or behavioral patterns could reveal deeper insights into specific user needs.
2. **Predictive Analytics**: Use machine learning to predict high-value user retention and churn rates, helping target users at risk of disengagement.
3. **Personalized User Experiences**: Leveraging insights on device preferences and transaction habits, the platform can offer tailored experiences for mobile vs. desktop users or region-specific features.
4. **Real-Time Monitoring**: Implementing a real-time dashboard to monitor user activity, session durations, and transaction volumes can help in dynamic decision-making and timely user engagement actions.